

Sentence Analyzer – a tool for learning Japanese



Anh Nguyen '18 | Advisor: Professor Takunari Miyazaki | Department of Computer Science, Trinity College

Problem

Japanese doesn't have spaces in-between words, which can make it hard for machines and people without a good understanding of the language to understand the grammar structure and the words that appear in a sentence.

Unrecognized Japanese name entities can be misinterpreted and mistranslated, leading to incorrect information extraction.

Project Introduction

Sentence Analyzer performs morphological analysis on Japanese texts with enhanced name-entity recognition ability & can display the analysis in Japanese/English. The improved accuracy in information extraction can be useful in many fields (translation, AI,..). This software also provides English translation of each word, which is convenient for those learning Japanese.

Analysis Algorithm

Process the initial analysis produced by MeCab through a corpus of 20,000 Japanese last names built from Myoji-Yurai to detect names

Detect other unrecognized names by running the analysis of all cases that a name can appear in a sentence, based on the placement of each word and its part-of speech tags

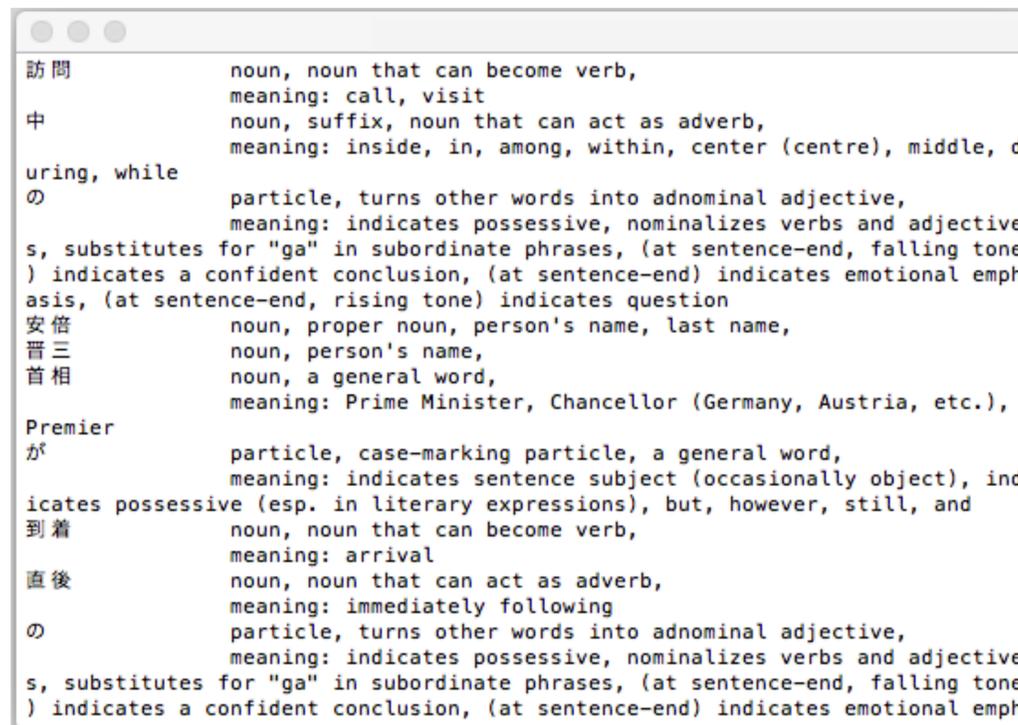
The whole analysis is translated into English using a original part-of-speech dictionary in the app, the meaning of each word can be displayed by processing the input through a simplified version of a widely used Japanese-English dictionary

Features

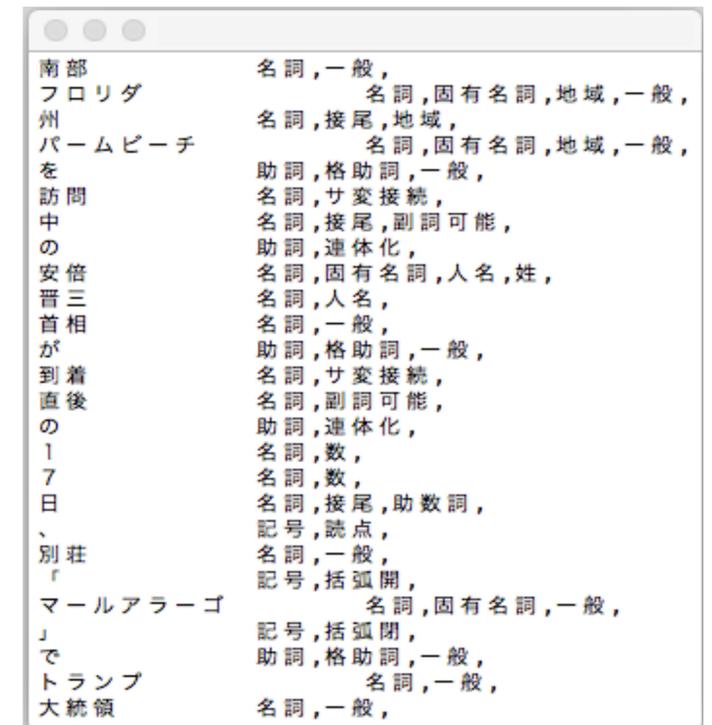
Perform morphological analysis on a file of Japanese texts:

- A list of all words and their part-of-speech tags - their roles in the grammatical structure of the sentences
- Enhanced name-recognition ability, especially for Japanese's last names and person's name
- English translation (if available) of each word in the original Japanese text input

The analysis can be displayed in either Japanese or English, which can be customized to each user's preference.

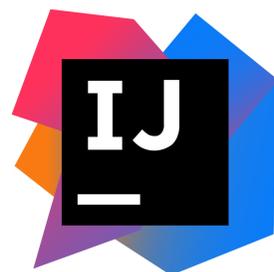


Analysis in English with translation of each word



Analysis in Japanese

Technology & Resources



MeCab



JMdict/EDICT Project



Acknowledgements

A special thanks to Professor Miyazaki and the Department of Computer Science for the guidance and support throughout this project.